Team Cloud Optimizer's - Project Proposal, Fall 2017

Members – Vinay Khedekar, Tejas Ghalsasi, Momtaz Afridi

1. Research Title–

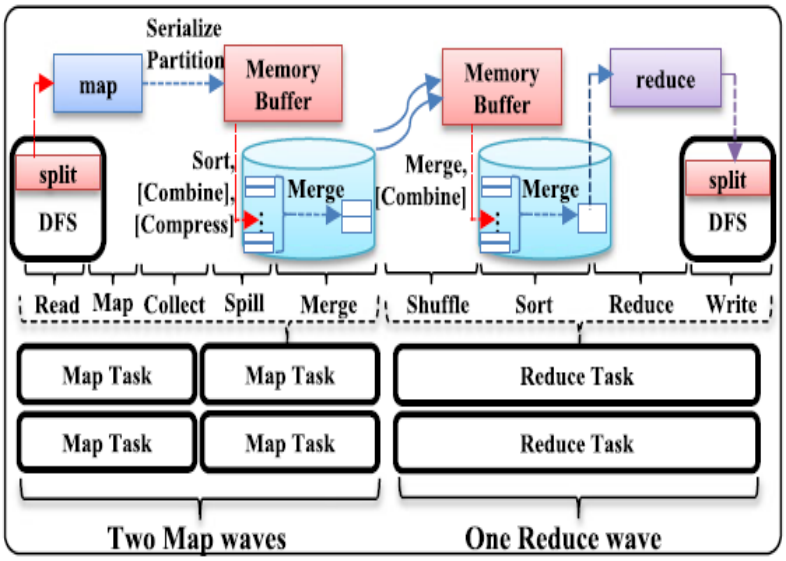
Research study – An OpenSimplex approach: Performance optimization techniques for Hadoop MapReduce system by optimizing Hadoop system configuration parameters.

2. Project Description –

The goal of this research project is to conduct wide research on Hadoop MapReduce system's configuration parameters, to develop new performance optimization techniques for Hadoop MapReduce system. As achieving optimal results from a Hadoop implementation begins with highly efficient system configurations and default values of these parameters do not result in satisfactory performance and therefore it is important to tune them. The research work mainly focuses on deep study of 190 configuration parameters of Hadoop system which includes study of parameters of Linux file system, parameters of Hadoop, Map/Reduce-Specific Configurations and Hadoop job parameter. Also, the goal of research study is to develop a dimensionality-free method which can automatically tune the configuration parameters by considering the cross-parameter dependencies. The new 'OpenSimplex approach' an optimization technique which will analyze the Hadoop system cross parameter configurations based on application setup and tune them to achieve desired system performance.

Project Setup Environment –

Setup of Hadoop MapReduce System and Configuration Parameters



Hadoop MapReduce System

Configuration Parameter
io.sort.factor
mapred.job.shuffle.merge.percent
mapred.output.compress
mapred.inmem.merge.threshold
mapred.reduce.tasks
io.sort.spill.percent
mapred.job.shuffle.input.buffer.percent
io.sort.record.percent
io.sort.mb
mapred.compress.map.output

Project Setup environment involves –

1. Setup of Hadoop MapReduce framework
2. Configurations of Linux file system
3. Configurations of Hadoop framework
4. Configurations of MapReduce
5. Hadoop Job parameters
6. Related many configurations

3. Project Deliverables –

1. Wide research on Hadoop MapReduce system's, 190 configuration parameters' and their cross-parameter dependencies to develop efficient optimization techniques.
2. OpenSimplex approach: Development of a dimensionality-free method which can automatically tune the configuration parameters by considering the cross-parameter dependencies.

4. Project Skills Needed –

- **Framework –** Hadoop MapReduce, Amazon Web Server (AWS)
- **Language & Scripts –** Java, TCL (Tool Command Line), Ubuntu Terminal
- **Operating System-** Linux, Ubuntu
- **Scripting –** Server Scripting, 4GL scripting
- **Server Configuration skills**